

Pronóstico de series temporales aplicada a la predicción de eventos climáticos severos para la Región del Valle de Catamarca

Pucheta, Julián¹, Salas, Carlos², Herrera, Martín², Arrieta, Fernando³, Patiño, H. Daniel⁴ y Rodríguez Rivero, Cristian⁵

(1) *Departamento de Electrónica, Facultad de Ciencias Exactas, Físicas y Naturales, Universidad Nacional de Córdoba, Av. Vélez Sarsfield 1611, Córdoba, Argentina. jpucheta@unc.edu.ar.*

(2) *Facultad de Tecnología y Ciencias Aplicadas, Universidad Nacional de Catamarca, Catamarca, Argentina. csalas@tecno.unca.edu.ar*

(3) *Universidad Austral, <https://www.austral.edu.ar>, Buenos Aires, Argentina. arrietafernando@gmail.com*

(4) *Instituto de Automática, Facultad de Ingeniería, Universidad Nacional de San Juan, Av. Libertador General San Martín (Oeste) N° 1109. San Juan, Argentina. dpatino@inaut.unsj.edu.ar*

(5) *Cardiff School of Technologies, Cardiff Metropolitan University, Western Avenue, Cardiff, CF5 2YB, Wales. CRodriguezRivero@cardiffmet.ac.uk*

RESUMEN

En la Región de la ciudad de San Fernando del Valle de Catamarca hay una serie de antecedentes de eventos climáticos severos que han quedado registradas por sus impactos negativos en la sociedad. Desde la disciplina de control automático se trata de proponer una solución al hecho de estar alerta y prevenir acciones que mitiguen los daños mediante un sistema de modelado de procesos dinámicos. Se han generado diversas soluciones al modelado de procesos climáticos partiendo de series temporales de valores de lluvia, y se proponen algoritmos de ajuste, validación y prueba para el caso particular del Valle de Catamarca. Se emplea la serie de Mackey Glass para realizar el diseño preliminar, y luego las series de lluvia obtenidas desde una central meteorológica. Se diseñaron algoritmos basados en modelos dinámicos lineales, ya que evidencian un buen desempeño para realizar extrapolaciones con simulación de Monte Carlo que emplea ruidos de distribución gaussiana, gamma y Poisson. Se muestran resultados de predicción para las series de Mackey Glass 17 y 30 con horizonte de predicción de 24 pasos adelante, y para predicción de series de lluvia con horizonte de predicción de cuatro días. Se detallan las simulaciones Monte Carlo en cada caso.

ABSTRACT

In the Region of San Fernando del Valle de Catamarca, there are various antecedents of severe climatic events that have been recorded due to their negative impacts on society. In this paper, an approach from the automatic control discipline to being aware of and foreseeing actions that mitigate the damage through a dynamic process modeling system is proposed. Several solutions for modeling climatic processes based on rainfall value time series have been generated, and algorithms for tuning, validating, and testing in the case of the Catamarca Valley have been proposed. The Mackey Glass series was used to carry out the preliminary design, and then the rain series obtained from a meteorological station was used. Algorithms based on linear dynamic models were designed because they show good performance in performing extrapolations with Monte Carlo simulation using Gaussian, Gamma, and Poisson distribution noises. The forecast results are shown for the Mackey Glass 17 and 30 time series, with a forecast horizon of 24 steps ahead. For the rainfall time series forecast, a prediction horizon of four days was used. The Monte Carlo simulations were detailed for each case.

Palabras Clave — pronóstico de series temporales- modelos dinámicos- predicción de series temporales- Simulación Monte Carlo.

Keywords — time series forecasting - dynamic models – prediction - Monte Carlo Simulation.

1. INTRODUCCIÓN

En la Región de la ciudad de San Fernando del Valle de Catamarca hay una serie de antecedentes de inclemencias de eventos climáticos que han quedado registradas por sus impactos en la sociedad, por ejemplo, en el año 2020 se detalla por el medio Catamarca Actual (2020) y el diario El Ancasti (2020). Además, existieron eventos en el año 2021 como registra el diario La Nación (2021) y Página 12 (2021) en la propia Ciudad. Estos eventos generan una demanda de un sistema de alerta que tenga capacidad de predecir el evento con una antelación útil para la toma de decisiones.

1.1 Presentación del problema

Para ello, se requiere de realizar sistemas de predicción y pronóstico enfocados en el evento de interés, con la historia de valores climáticos que dan información para proponer un sistema con exactitud adecuada a la cantidad de tiempo previo que se requiera en los cuales ya se tienen diversos resultados que sirven como referencia Pucheta *et al.* 2012, Rodríguez Rivero *et al.* 2017a, Rodríguez Rivero *et al.* 2017b y Rodríguez Rivero *et al.* 2017c.

1.2 Análisis exploratorio de posibles soluciones

Se cuenta con datos de sensores remotos, como es el caso del proyecto Prediction Of Worldwide Energy Resources (POWER) de NASA, y puede emplearse para hacer un análisis preliminar del comportamiento de las lluvias a lo largo de los años. El proyecto POWER se inició para mejorar el conjunto de datos de energía renovable actual y crear nuevos conjuntos de datos a partir de nuevos sistemas satelitales. El proyecto POWER se dirige a tres comunidades de usuarios, los que se enfocan en la Energía renovable, o los que estudian Edificios sostenibles y los de Agroclimatología. La temática del presente trabajo se enmarca en los últimos dos usuarios. Esta fuente de datos está activa desde 1981, tiene una resolución de 50Km por 50Km y una latencia de 48Hs. Este dato es importante para la generación de una alerta por evento severo ya que, en el momento actual, el dato más nuevo corresponde a dos días atrás y la dinámica de los eventos son más rápidos. En la Fig. 1 se muestran los últimos años de datos de lluvia diaria, con un

dato por día y una latencia de dos días. Nótese que, con esta latencia, para mostrar el dato de lluvia actual, hay que predecir dos días y si además se trata de generar una alerta para tomas de decisiones, este umbral aumenta la incertidumbre en la información generada por el sistema requerido.

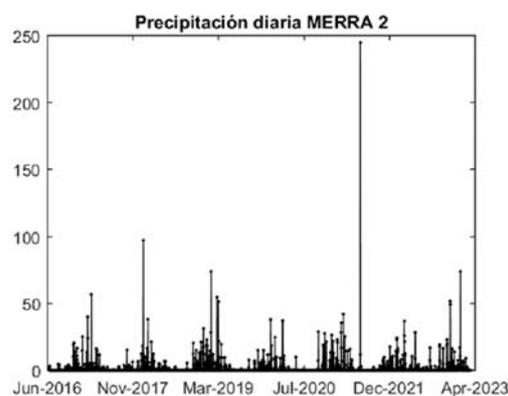


Figura 1. Visualización de las series temporales de lluvia obtenidas desde NASA POWER. Los datos tienen una latencia de 48hs.

Por lo tanto, además de emplear los datos provenientes de sensores remotos para un análisis de largo plazo (desde 1981 a la fecha), se está desarrollando y manteniendo un sistema de medición mediante instrumental a través del Gobierno de la Provincia de Catamarca (Central CAPE, 2016) con la interfaz que se muestra en la Fig. 4. Estos datos tienen una latencia de 15 minutos, pero el sistema está expuesto a las inclemencias de las tormentas por lo que en momento exacto del evento tiene grandes posibilidades de no funcionar. Aquí se puede destacar la necesidad del sistema de predicción con un tiempo de uno o dos días para la toma de decisiones. En la Fig. 2 se muestran los datos obtenidos desde la central de adquisición que tiene el Gobierno de Catamarca. El objetivo se puede sintetizar en generar un valor probable de precipitación para los dos días siguientes al actual. En este trabajo se genera un pronóstico para los cuatro días siguientes, con fines de facilitar el análisis. Para el caso de las MG el horizonte es de 24 días adelante.

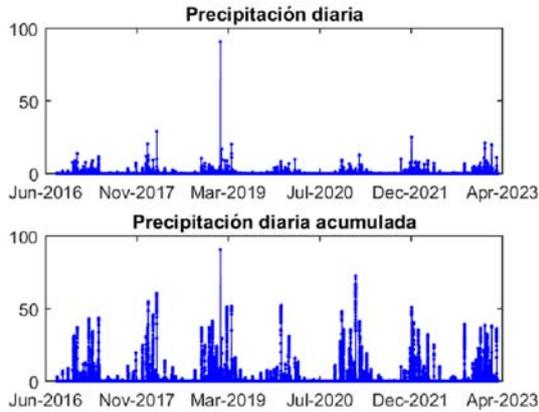


Figura 2. Visualización de las series temporales de lluvia obtenidas desde la Central CAPE. Arriba se detallan los valores de precipitación disponibles cada 15 minutos, y abajo los valores acumulados diarios.

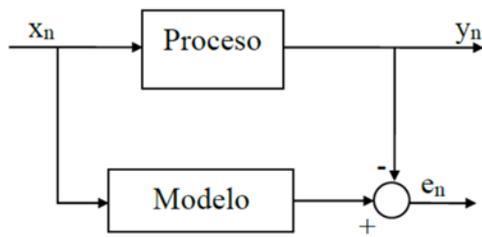


Figura 3. Representación de la señal de error entre un proceso real y su modelo. El modelo tiene parámetros que son ajustados mediante la secuencia de error.

2. MATERIALES Y MÉTODOS

Se propone realizar un modelado que toma las entradas y salidas para el cálculo como se muestra en la Fig. 3. El modelo contiene parámetros que deben sintonizarse. Para ello se propone el uso del ajuste por el gradiente descendente, donde el error de ajuste está dado por

$$e_n = y_n - f(\theta_n, x_n, y_{n-1}) = y_n - \hat{y}_n \quad (1)$$

donde y_n es el valor actual de la salida del proceso, $f(\cdot)$ indica la salida del modelo que se está ajustando con parámetros θ_n a definir en cantidad y en valores, y x_n es el valor de la entrada actual. Se asume que se dispone de un total de datos muestreados con intervalos unitarios desde 0 hasta n . Aquí $f(\cdot)$ se implementa

como un sistema dinámico lineal, definiendo al vector de los parámetros θ como

$$\theta^T = [a_2 \quad a_3 \quad \dots \quad a_{o_y} \quad b_1 \quad b_2 \quad \dots \quad b_{o_x}] \quad (2)$$

se define que la función $f(\cdot)$ es

$$\begin{aligned} \hat{y}_n(\theta, x_n, y_{n-1}) = & b_1 x_n + b_2 x_{n-1} + \\ & + b_{o_x} x_{n-o_x+1} \\ & - a_2 y_{n-1} - \dots - a_{o_y} y_{n-o_y+1} \end{aligned} \quad (3)$$

donde los parámetros serán ajustados en el sentido de minimizar el error (1) en donde interviene la medición actual y_n , que no aparece en (3). Para el ajuste de dichos parámetros, se define el funcional a minimizar respecto de los parámetros como

$$J_n = \frac{1}{2} \sum_{n=n_i}^{n_f} e_n^2 = \frac{1}{2} \sum_{n=n_i}^{n_f} \{y_n - f(\theta, x_n, y_{n-1})\}^2 \quad (4)$$

donde están definidos dos tiempos específicos de ajuste que son n_i y n_f como tiempo inicial y tiempo final para el ajuste del modelo. Este ajuste se realiza una vez por cada iteración, durante la que se hará variar cada componente de θ . Los parámetros serán ajustados durante una etapa de tiempo, pero luego quedarán fijos para generar la predicción.

Para ajustar los parámetros del modelo, que son las componentes a_i y b_i del vector θ , se procede empleando el método del gradiente descendente. Para ello, se empieza minimizando a (4) respecto de dichas componentes hallando sus derivadas parciales, por lo tanto

$$\frac{\partial J_n}{\partial a_i} = -e_n y_{n-i+1}, i = 2, 3, \dots, o_y. \quad (5)$$

$$\frac{\partial J_n}{\partial b_i} = e_n x_{n-i+1}, i = 1, 2, \dots, o_x. \quad (6)$$

Con estos incrementos se definen las cantidades del algoritmo de gradiente descendente para el instante n , como

$$b_i := b_i + \gamma e_n x_{n-i+1}, i = 1, 2, \dots, o_x. \quad (7)$$

$$a_i := a_i - \gamma e_n y_{n-i+1}, i = 2, 3, \dots, o_y. \quad (8)$$

con el valor de γ como ganancia de ajuste o paso de ajuste. Con las variables $\{x, y\}$ correctamente condicionadas, éste valor puede estar en el orden de 10^{-3} a 10^{-6} para los casos aquí estudiados. Por

lo tanto, dados un conjunto de datos de mediciones hasta un instante n , $\{x_n, y_n\}$ se requiere elegir la relación de correspondencia dinámica entre la entrada x y la salida y de tal manera que el modelo muestre máxima evidencia. El algoritmo expresado en (7)-(8) depende de las condiciones iniciales de θ , ya que es incremental, por lo que se proponen valores de orden reducido en dimensión y luego se procede a aumentar los valores de σ_x , σ_y si el modelo no muestra la suficiente evidencia.

La predicción de los valores siguientes al instante n estarán dados por

$$\hat{y}_{t+D} = [-\hat{y}_{t+D-1} \cdots -\hat{y}_{t+D-n} \ u_{t+D-1} \cdots \ u_{t+D-1-m}] \cdot \theta = \varphi(t+D)^T \cdot \theta. \quad (9)$$

En la predicción sí aparece la señal de entrada u que tendrá una distribución Gaussiana, Gamma o de Poisson y se genera una simulación Monte Carlo.

El criterio para medir la evidencia del modelo con el vector θ actual consiste en estudiar el comportamiento la autocorrelación detallada por Oppenheim *et al.* (1999). Así, el modelo obtenido que presente la función de autocorrelación más parecida a la que presentan los datos originales, será el modelo elegido.

Se supone disponible una determinada cantidad de datos observados, por lo que existirán errores debido a que la cantidad es finita. La función de autocorrelación para sistemas ergódicos se define como

$$\phi(\tau) = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T x(t)x(t+\tau)dt \quad (10)$$

y para intervalos discretos unitarios se tiene

$$\phi(j) = \frac{1}{N-j} \sum_{i=0}^{N-j} x(i)x(i+j) \quad (11)$$

con $j=0, 1, \dots, M$, y además $i=0, 1, \dots, N$. Donde M es el número de puntos de la función autocorrelación, N es el número de puntos medidos de la señal. Cabe aclarar que normalmente M es al menos la mitad de N , ya que la función intercorrelación está definida para N mucho mayor que M para asegurar una buena representación. Esto se ve en la definición (10), donde la integral va desde $-\infty$ a ∞ donde deberían

estar definidas las señales y en la práctica se trabajan con valores discretos limitados.



Figura 4. Sistema de visualización de variables ambientales disponible en la web vía la Central CAPE (2016).

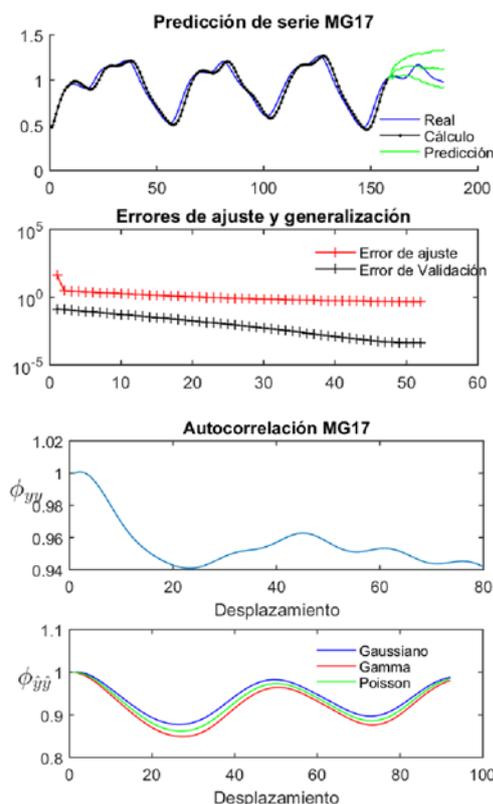


Figura 5. Ajuste del algoritmo para correr el Monte Carlo con ruido Gaussiano, Gamma y Poisson. Para el primero fue de media nula con desviación de 0,31, para el Gamma con un factor de forma 0,03 y para Poisson con un λ de 0,08.

3. RESULTADOS

Se implementa el algoritmo detallado para probarlo en series de dato Mackey Glass 17 y 30

siguiendo el método de modelado de series de lluvia de (Rodríguez Rivero *et al.* 2017c) y luego en las series de lluvia de la Región del Valle de Catamarca de la Central CAPE.

Se aplicó el método descrito y se obtuvieron los resultados de predicción que se muestran en las Fig. 5 para la serie MG17, Fig. 7 para las MG30, y en las Fig. 6 y Fig. 8 para el caso de las series de lluvia. Para estas últimas se ha agregado una figura extra debido a la necesidad de ver los detalles. En todos los casos, se muestran las firmas de autorrelación a la derecha, con las funciones superpuestas.

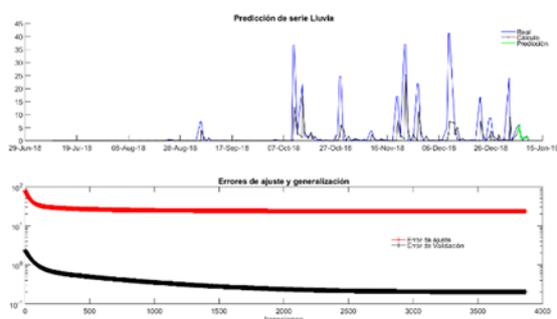


Figura 6. Visualización de la predicción las series temporales de lluvia obtenidas desde Central CAPE (2016).

4. CONCLUSIONES

Se ha mostrado un método para pronóstico de series temporales que permite generar la predicción de eventos climáticos severos con un tiempo de antelación de cuatro días. Así planteado, se tiene que realizar el ajuste del algoritmo hasta hallar la máxima evidencia del modelo mediante las gráficas de autocorrelación. Este procedimiento es consumidor de tiempo de cálculo, pero se espera lograr minimizar este costo mediante técnicas de aprendizaje automático, como en Pucheta *et al.* 2012. Este método asume a los datos con errores acotados y que no hay faltantes. El método de evidencia de modelo empleado aquí es más simple que el usado previamente en casos similares, que requieren de grandes volúmenes de cómputo.

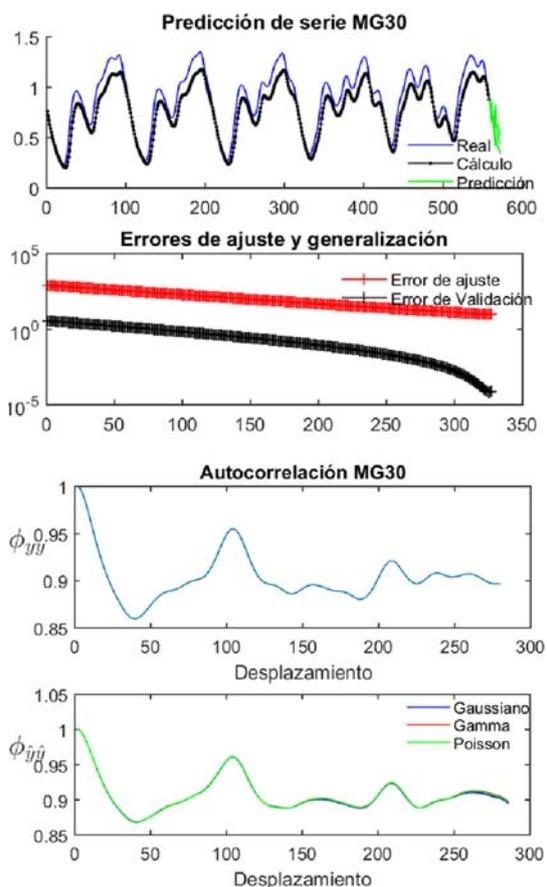


Figura 7. Ajuste del algoritmo para correr el Monte Carlo con ruido Gaussiano, Gamma y Poisson. Para el primero fue de media nula con desviación de 0,44, para el Gamma con un factor de forma 0,03 y para Poisson con un λ de 0,08.

5. AGRADECIMIENTOS

Los autores agradecen a la Universidad Nacional de Catamarca, al Ministerio de Ciencia e Innovación Tecnológica del Gobierno de Catamarca, a la Universidad Nacional de Córdoba, a la Universidad Nacional de San Juan y a la Cardiff Metropolitan University por su apoyo en el desarrollo de proyectos involucrados en el Control y modelado de procesos de dinámica lenta, media y rápida.